*Acta Psychiatr Scand 2011: 124: 70–72*

DOI: 10.1111/j.1600-0447.2011.01708.x


*ACTA PSYCHIATRICA SCANDINAVICA*

# Debate section

## Discussion paper

# Psychiatry and the scientific fallacy

For Richard Feynman, Nobel Physics Prize, there was one specific feature of 'true science':

> *[…]the idea that we all hope you have learned in studying science in school – we never explicitly say what this is, but just hope that you catch on by all the examples of scientific investigation […] It's a kind of scientific integrity, a principle of scientific thought that corresponds to a kind of utter honesty – a kind of leaning over backwards (1).*

Evidence-based medicine (EBM) is now a science. So is psychiatry: DSM-III emphasized diagnostic reliability as the condition for scientific psychiatry; randomized controlled trials (RCTs) and meta-analyses are the gold standard to measure therapeutic effects; psychopathology evolved into a complex statistical discipline; psychiatric disorders are investigated by neuroscientists; and, since the 1990s, psychiatrists use 'new-generation' antidepressants and antipsychotics. A scientific nosology, complex tools, rigorous statistics and modern, sophisticated, receptor-specific drugs – this is the current *Zeitgeist*: psychiatry is finally leaving the Freudian wars of the past and joining up with science.

### What are the facts?

At the end of the 1990s, the National Institute of Mental Health sponsored three large EBM studies, CATIE, STAR*D, and STEP-BD, aimed at updating treatment guidelines for schizophrenia, depression, and bipolar disorder, because published RCTs did not compare the new drugs. As the *external validity* of RCTs was questioned, these *effectiveness* studies tried to emulate 'real-world' practice: wide inclusion, 'long-term' follow-up, and patient's participation in treatment (2).

i) CATIE: schizophrenic subjects were randomized between various atypical antipsychotics and pherphenazine; those interrupting treatment were randomized between atypicals and clozapine if previous treatment was inefficient, or another atypical if previous treatment was not tolerated;
ii) STAR*D: after insufficient response to citalopram, depressive subjects went through various randomized strategies (switch/augmentation with/without psychotherapy); 2-3 more steps were planned for non-remitters;
iii) STEP-BD: bipolar subjects received 'best-practice treatment' and entered various RCTs, e.g. treatment for bipolar depression.

Therapeutic results were poor:

i) In CATIE (3), 74% interrupted treatment over 8 months; new drugs and pherphenazine were similarly (in)effective : symptom scores decreased by only 2–10%. Clozapine was slightly better.
ii) In STAR*D (4), 65% never entered *sustained* remission (i.e. >12 months): 30% resisted all treatments, 50% of the remitters relapsed, despite escalating doses, switches, etc.; the various strategies were similar in their (in)effectiveness;
iii) In STEP-BD (5), 70% never entered 'sustained' (8 weeks!) remission over 2 years. Antidepressants proved inefficient in bipolar depression. Lamotrigine was somewhat effective in resistant depression – but later studies have been negative; inositol and risperidone were not. The 'randomized relapse prevention trial' and the observational 'best-practice pathways' (Acute Depression, Refractory Depression, Acute Mania, Refractory Mania, Rapid Cycling, Relapse Prevention, Pregnancy, Substance Abuse, and Other Comorbidity) were never published – intriguing, as these are the crucial clinical questions.

The authors concluded that new antipsychotics are not better than older ones but more expensive; no individual factor predicts antidepressant response; one should 'monitor' depressive patients, adapt their treatment, and follow them after remission, although STAR*D showed low response despite 'measurement-based' treatment, high relapse rate despite 'diligent follow-up', and did not address relapse treatment; the sickest patients (comorbid, chronic, highly symptomatic, etc.) are the sickest patients; bipolar depression is severe and difficult to treat.

### Keep your enthusiasm

In other words: truisms, poor treatment responses, and no answer to the original questions. The trials produced interesting epidemiological data but no evidence to guide evidence-based treatments, which was their stated goal. The results, however, stirred up enthusiastic comments, in editorials and reviews or from the authors themselves, sometimes even concerning the drugs' effectiveness, especially for STAR*D, which seems truly inspiring:

i) CATIE: '*The results could be viewed as discouraging […]. The value of CATIE is that it provides solid evidence to help clinicians and their patients make the difficult decisions needed to optimize the treatment of schizophrenia with the compounds currently available*' (6); CATIE provided some discouraging evidence about old and recent antipsychotics similar (in)efficiency, and no evidence to guide the choice among them.
ii) STAR*D: '*These findings are encouraging, both in the justification for attempting to induce a remission among patients for whom an initial antidepressant trial has failed and in their demonstration that multiple agents may achieve such a remission […] Affective neuroscience and effectiveness trials such as STAR*D should help us*

*identify new targets for treatment and patients for whom the treatments will be the most effective and best tolerated*' (7); STAR*D just failed to do so, and results were very discouraging about the possibility of sustained remission.

'*STAR*D […] has provided important evidence applicable to clinical decision making by ordering treatments by relative efficacy or tolerability at specific therapeutic junctures. Future studies must build on these trials […]*' (8); treatments were equivalent and not ordered in any way.

'*The STAR*D trial provides robust, real-world data that can be applied broadly to both primary and specialty care settings. The study confirms that different people respond to different treatment strategies […] The STAR*D team concluded that future research should be targeted to identify the best multi-step treatment options for individuals, especially those with treatment-resistant depression*' (9); this was precisely STAR*D team's question: they conclude that they should start over again.

iii) STEP-BD: '*The results of the STEP-BD study argue strongly that genetic, brain imaging, and neurobiological studies of bipolar disorder must be accelerated to help define who will respond best to which treatments in the long term*' (10); these results argue as strongly that genetic, brain imaging, and neurobiological studies of bipolar disorder so far have proved totally useless for treatment.

'The Clinical Trial That Keeps on Giving', commented *Psychiatric News* on a *Psychiatric Services* issue devoted to STAR*D: 'the data and conclusions from this large trial continue to provide valuable insight to clinicians, health care policymakers, and researchers'. *Psychiatric Services* was very positive about STAR*D and actually claimed that a major problem is that clinicians do not apply STAR*D therapeutic strategies (11):

> *The continuing gap between knowledge and practice is one of the most vexing problems facing our health care system. The 17-year latency period before consistent application of new knowledge to ordinary practice likely proves fatal for thousands of people each year. The gap stems from persistent problems in the training and support of clinicians as well as in the organization and financing of services. Addressing these problems will be a core challenge in efforts to reform health care. It is critical that this gap be closed.*

In the end, even AJ Rush, STAR*D main investigator, hints that what matters is not the evidence but the journey toward evidence (12):

> *Finally, on a personal note, large efforts like STAR*D are the ultimate exercise in ''delayed gratification''. But at the end of the day, the journey—the process of working with outstanding investigators and committed staff and patients—is its own unique reward. No single trial can answer more than a few specific questions, but such efforts can develop new clinical or research methods and raise important questions for further study.*

Badly needed Taoist wisdom, as STAR*D did not answer the questions asked.

## Publication bias: something rotten in the land of EBM?

Recently, a meta-analysis on antidepressants found a lower efficacy when unpublished studies were included (13). It stimulated much discussion on antidepressant efficacy and the influence of the pharmaceutical industry, as it demonstrated the magnitude of the *publication bias*, which was obviously underestimated, meta-analysis of *published* studies being the EBM gold standard of proof.

One suggested cause was the influence of pharmaceutical industry. In 2005, the International Committee of Medical Journal Editors required prior registration in a public trials registry before considering the publication of interventional trials, so that unpublished data would be accessible. Turner strongly illustrated the importance to 'lean over backward': pharmacological meta-analyses now mostly try to include unpublished data. However, the publication bias is not limited to interventional research, and many meta-analyses are still based on published studies, including older RCTs, addressing non-interventional questions, or simply not including unpublished material. The poor *external validity* of RCTs using highly selected subjects – one of the main incentives for CATIE et al. – is also a major problem. Still, meta-analyses determine treatment guidelines, insurance policies, putative diagnostic categories, and health care policies, without much criticism.

## DSM: too big to fail?

In a 2009 issue of the *American Journal of Psychiatry*, Darrel Regier et al. (14) discuss some novelties of the future DSM-5: dimensional criteria, consideration of gender and cultural context, an impairment dimension independent of diagnosis. These developments are reminiscent of early DSM-III criticisms: the categorical approach, lack of attention to context, and definition of 'mental disorder' were hotly debated in their time.

The article underlines that changes are necessary because of the high comorbidity rate, which suggests that current categories are invalid, and the fact that 'dimensional measures may provide better phenotypic expressions for linkage' to neuroimaging and genetic substrates. Changes are possible because 'original Robins and Guze validators have not confirmed the wisdom of the current structure', which means that DSM-III and IV have not been validated along the principles that guided their writing. The implication is that they were not validated because their structure is wrong.

Since 1980, research, health care policies, civil and criminal responsibility, insurance and benefit procedures have been based on DSM-III and IV categories. If those are invalid, one would expect some epistemological discussion. For instance, are Robins and Guze's criteria valid? If yes, what are their limits ? How were they actually applied? Is there a problem with the method that was used to apply them? Or is it just a matter of time, 30 years being too short a period, and Regier et al. are just too impatient – as Spitzer seems to argue ?

We are not saying that DSM-III and IV are wrong, nor that Spitzer was wrong, but if Regier et al. even only allude to that, they must question the Spitzer/Washington University approach. This is not the case – they speak of DSM-III most enthusiastically:

> *After almost forty years of testing these hypotheses* [Robins and Guze's hypotheses: specific criteria will establish the scientific validity of the proposed

diagnostic categories], *we are impressed by the remarkable advance in research and clinical practice that were facilitated by having explicit criteria that produced greater reliability in diagnosis across clinicians and research investigators in many countries*.

How could invalid categories facilitate such 'remarkable advance', which is, by the way, supported by no evidence? We are far from Feynman's 'utter honesty' and 'leaning over backward'.

## A gap to be closed?

In view of the disappointing results of CATIE et al. and the mentioned methodological difficulties (problematic diagnostic categories, publication and subjects' selection biases), one would expect some crisis in psychiatric research. But business runs as usual: journals keep publishing positive RCTs with highly selected subjects diagnosed with DSM-IV criteria, and meta-analyses are scrutinized for some insignificant advantage of one or another antidepressant (15). Not that research should be stopped until every epistemological issue has been solved, but some debate, minimal modesty and self-criticism would be welcome when results are discussed.

Spitzer hoped for a scientific confirmation of DSM-III categories that did not take place. Logical empiricism did not validate our nosology, which did not help developing new concepts for psychiatric diseases or new treatments. Maybe psychiatric diseases are very complex and it takes a long time to break through; maybe the chosen methods of investigation are inadequate. The unwavering enthusiasm for modern psychiatry and its 'remarkable advance' is simply not scientific. Here is the true 'gap', between scientists' enthusiasm and clinicians' disillusion. This is why CATIE et al. had so little impact: clinicians continue to prescribe the same treatments, knowing the 'evidence' is poor. A sign that there might be a real problem in psychiatric research that needs to be addressed, instead of concluding that 'more studies are needed'.

*Michael Saraga*[1], *Friedrich Stiefel*[2]
[1]*Associate Psychiatrist, Liaison Psychiatry, Department of Ambulatory Care and Community Medicine, University Hospital of Lausanne, Policlinique Médicale Universitaire, rue du Bugnon 44, 1011 Lausanne, Switzerland and* [2]*Head, Psychiatric Liaison Service, Department of Psychiatry, University Hospital of Lausanne, Service de Psychiatrie de Liaison, rue du Bugnon 44, 1011 Lausanne, Switzerland*
*E-mail: michael.saraga@chuv.ch*

## References

1. Feynman RP. ''Surely you're joking, Mr. Feynman!'': adventures of a curious character. 1997.
2. Landmark Council Session Spotlights ''Real World'' Trials. 2006 Available from: http://www.nimh.nih.gov/science-news/2006/landmark-council-session-spotlights-real-world-trials.shtml [updated 2006; cited 2010 September 16].
3. Lieberman JA, Strop TS, Mcevoy JP et al. Effectiveness of antipsychotic drugs in patients with chronic schizophrenia. N Engl J Med 2005;**353**:1209–1223.
4. Rush AJ, Trivedi MH, Wisniewski SR et al. Acute and longer-term outcomes in depressed outpatients requiring one or several treatment steps: a STAR*D report. Am J Psychiatry 2006;**163**:1905–1917.
5. Perlis RH, Ostacher MJ, Patel JK et al. Predictors of recurrence in bipolar disorder: primary outcomes from the Systematic Treatment Enhancement Program for Bipolar Disorder (STEP-BD). Am J Psychiatry 2006;**163**:217–224.
6. Freedman R. The choice of antipsychotic drugs for schizophrenia. N Engl J Med 2005;**353**:1286–1288.
7. Rubinow DR. Treatment strategies after SSRI failure–good news and bad news. N Engl J Med 2006;**354**:1305–1307.
8. Valenstein M. Keeping our eyes on STAR*D. Am J Psychiatry 2006;**163**:1484–1486.
9. Odds of Beating Depression Diminish as Additional Treatment Strategies are Needed. 2006 [updated 2006; cited 2010 February 17]; Available from: http://www.nimh.nih.gov/science-news/2006/odds-of-beating-depression-diminish-as-additional-treatment-strategies-are-needed.shtml.
10. Depaulo JR Jr. Bipolar disorder treatment: an evidence-based reality check. Am J Psychiatry 2006;**163**:175–176.
11. Shern DL, Moran H. STAR*D: helping to close the gap between science and practice. Psychiatr Serv 2009;**60**:1458–1459.
12. Rush AJ. STAR*D: what have we learned? Am J Psychiatry 2007;**164**:201–204.
13. Turner EH, Matthews AM, Linardatos E, Tell RA, Rosenthal R. Selective publication of antidepressant trials and its influence on apparent efficacy. N Engl J Med 2008;**358**:252–260.
14. Regier DA, Narrow WE, Kuhl EA, Kupfer DJ. The conceptual development of DSM-V. Am J Psychiatry 2009;**166**:645–650.
15. Parikh SV. Antidepressants are not all created equal. Lancet 2009;**373**:700–701.

Editor's note:
In accordance with the policy of Acta Psychiatrica Scandinavica, the authors of the present paper informed the authors directly quoted in their paper. Two responded and Dr Sagar V. Parikh wrote a reply which is published in connection with the above paper.